EPA
United States
Environmental Protection
Agency

# High Throughput Screening of Toxicity Pathways Perturbed by Environmental Chemicals


**Office of Research and Development**
National Center for Computational Toxicology

United States Environmental Protection Agency

# Outline

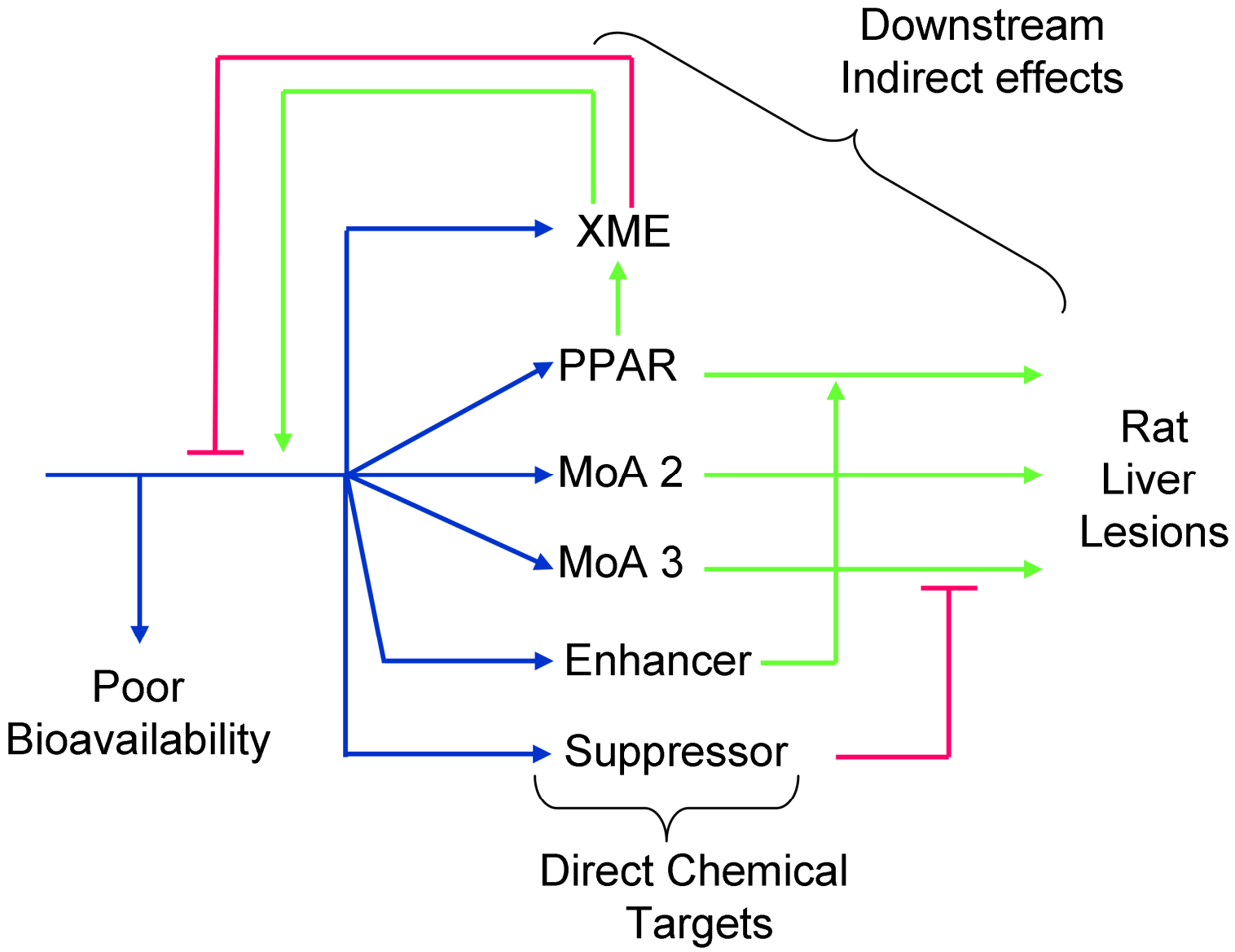
- Things to watch for in the data
- 3 Laws of Predictive Toxicology
- Rat liver tumor associations and signatures
- Future directions

United States Environmental Protection Agency

# Things to Watch for in the Data

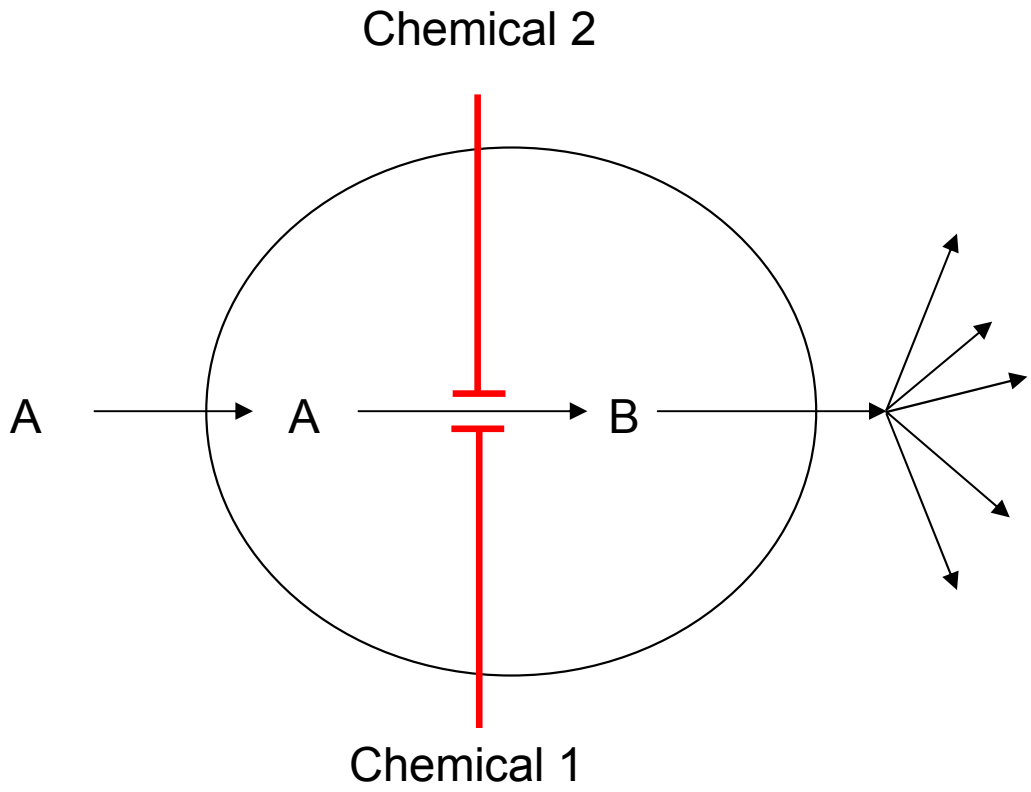
- ***in vitro* data**
  - What does the assay measure?
  - How reproducible are the assay results?
  - How do you compare an AC50 from one data set to another?
- ***in vivo* data**
  - What is the difference between missing data and a no call?
  - When is a no call really missing data?
  - How many mechanisms could give rise to a type of toxicity?
- **Chemicals**
  - What do the replicates mean?
  - Are “replicates” always the same chemical?
  - Consider metabolism / biotransformation

# Toxicity is Multi-factorial



# First "Law" of Predictive Toxicology
# "Direct Interactions"

If 2 chemicals have the same set of direct interactions and have the same bioavailability, they will cause similar toxicities

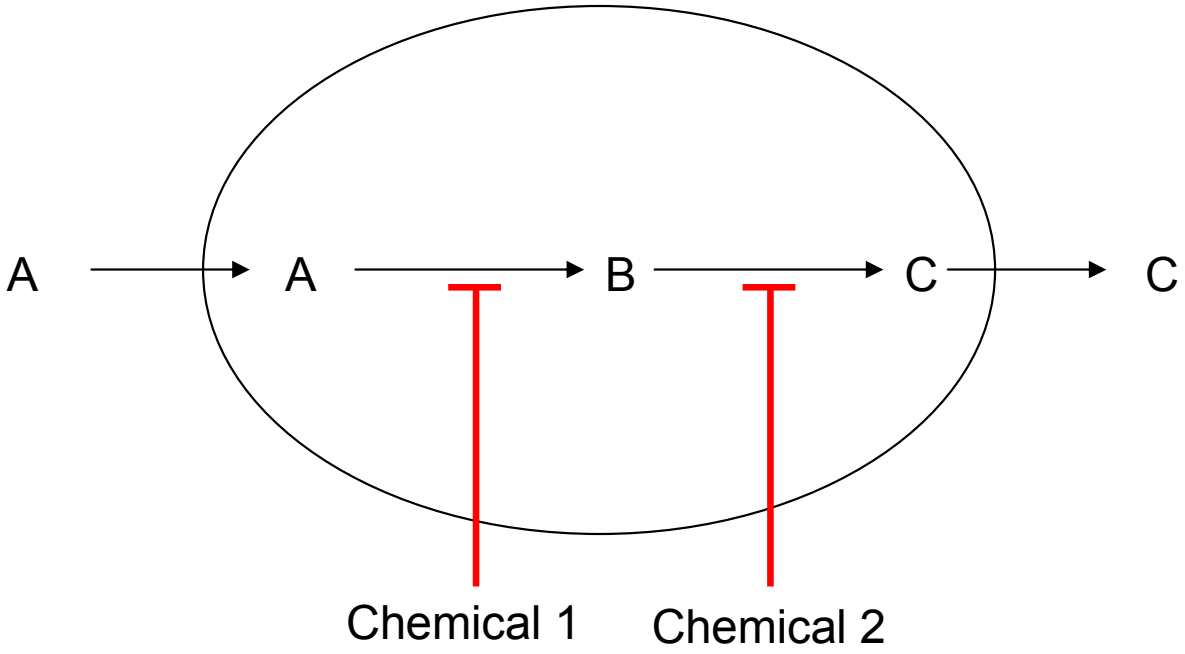



Relevant to discussion of:

- Adaptive Biology
- Differing Time Scales

# Second "Law" of Predictive Toxicology "Pathway Perturbations"

If 2 chemicals cause the same set of nearby indirect responses and have the same bioavailability, they will cause similar toxicities




Relevant to discussion of:

- Chemicals with different targets causing same downstream effects

# 3 "Laws" of Predictive Toxicology

1. If 2 chemicals have the same set of direct interactions and have the same bioavailability, they will cause similar toxicities

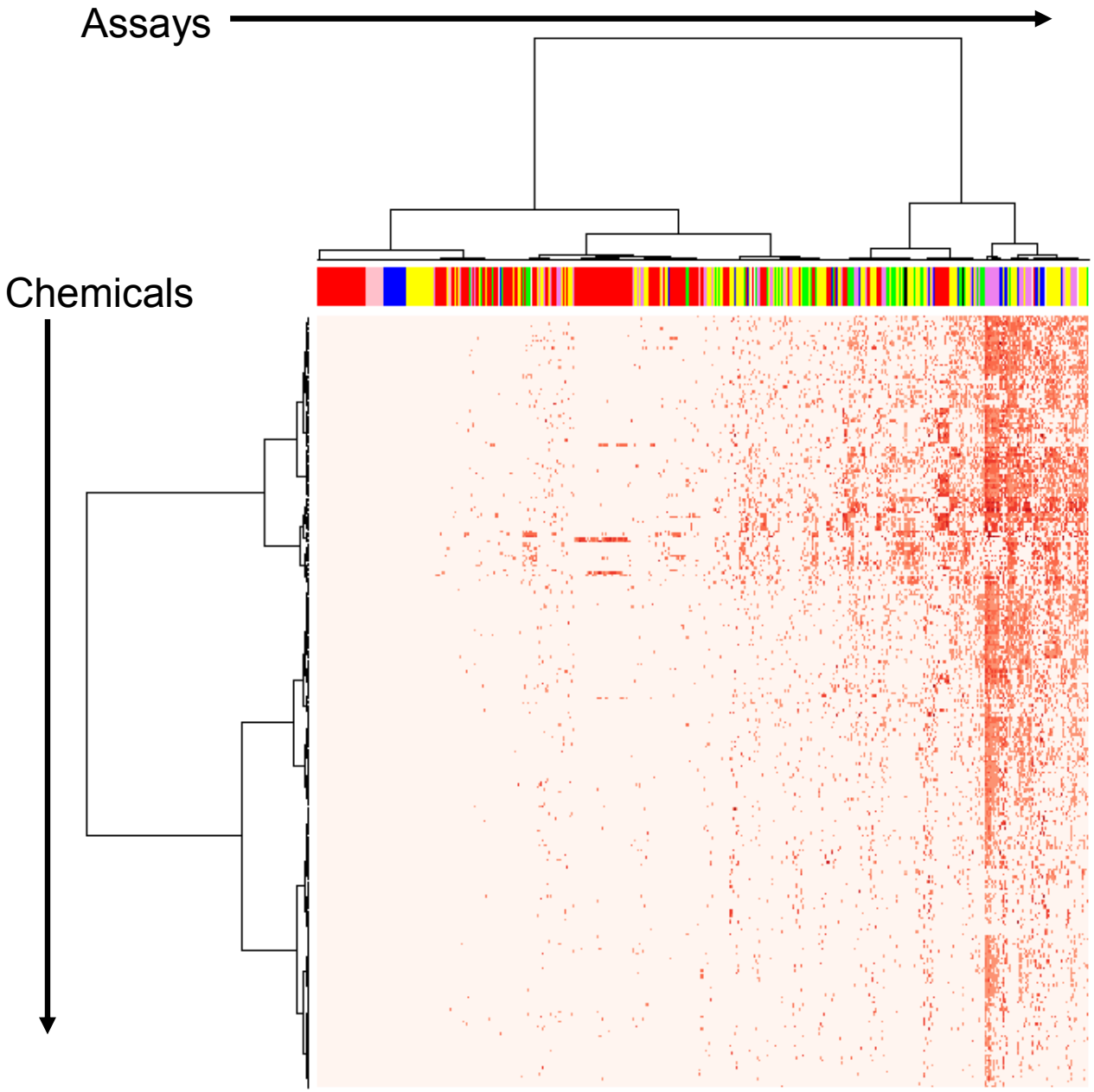
2. If 2 chemicals cause the same set of nearby indirect responses and have the same bioavailability, they will cause similar toxicities

3. It will be rare to find pairs of chemicals that behave so closely that Laws 1 and 2 can be applied with 100% confidence

Nonetheless …
(1) and (2) are the conceptual basis of using *in vitro* screening

(3) Is why we focus on prioritization before prediction

# The ToxCast In Vitro Data Set




Many hits – median value=50

Fewer in cell-free HTS (Novascreen, red)

Many hits are at or near top of tested concentration range

A few are in nanomolar range

United States Environmental Protection Agency

# Some Expected Hits from the Data

- Estrogen receptor (ER)
  - Bisphenol A, Methoxychlor, HPTE
- Androgen Receptor (AR)
  - Vinclozolin, Linuron, Prochloraz
- PPAR
  - PFOA, PFOS, Diethylhexyl Phthalate, Lactofen
- Mitochondrial Poisons
  - Azoxystrobin, Fluoxastrobin, Pyraclostrobin
- Acetylcholinesterase Inhibition
  - Multiple organophosphorus pesticides

United States Environmental Protection Agency

# Focus on Rodent Liver Toxicity

- Several liver lesion classes for rat, similar for mouse
  - All chemicals: 248
  - No Liver lesions:122
  - Any Lesion:126
  - Pre-neoplastic or neoplastic:58
  - Neoplastic: 21
  - Liver Proliferative Lesions:61
  - Liver Tumors (=neoplastic): 21

# Calculate Univariate Associations with Rat Liver Proliferative Lesions

- Significance Tests:
  - T-test (treat *in vitro* as continuous)
  - Chi-squared (treat *in vitro* as dichotomous, using 100μM as the cutoff)
- Significant associations are:
  - PPARA
  - PPARG
  - HMGCS2 (regulated by PPAR)
  - RXRA (dimerizes with PPAR)
  - CCL2
  - CCL26

# PPAR signaling and Rodent Liver Tumors

- PPAR is involved with lipid and fatty acid metabolism
- Xenobiotics can activate PPAR
  - Leads to peroxisome proliferation and hepatocyte hypertrophy

- PPAR-driven liver tumorigenesis does not seem to act in humans
  - But PPAR-driven hepatotoxicity is of concern (FDA)
  - PPAR is a target for human drugs to treat metabolic syndrome / diabetes

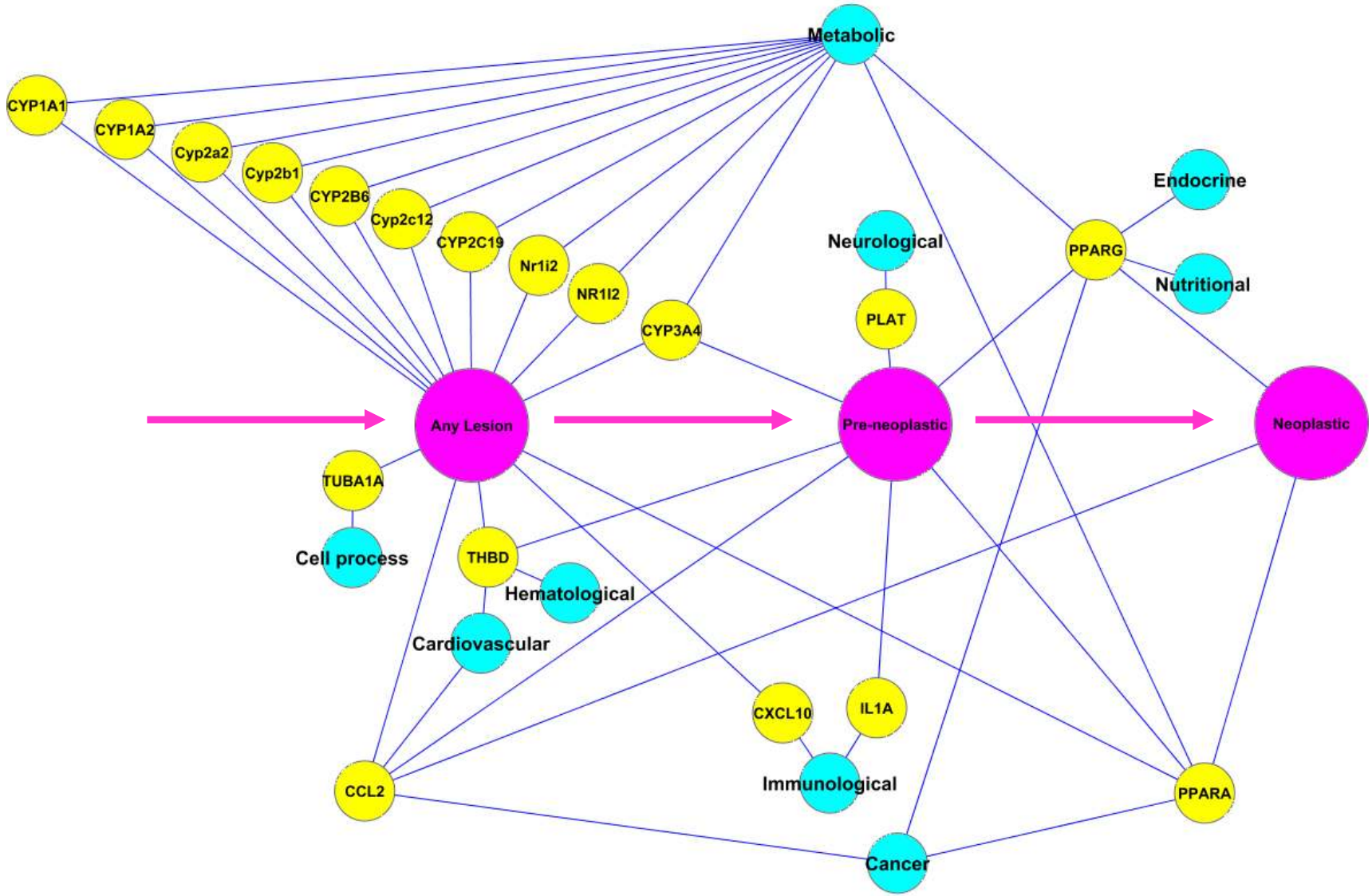
- 3 isoforms
  - PPARA / PPAR$\alpha$
  - PPARG / PPAR$\gamma$
  - PPARD / PPAR$\delta$

# CCL2 Associations with Environmental Chemicals and Liver Toxicity are Novel

- Chemokine (C-C motif) ligand 2
- Drives angiogenesis and tumor cell invasion
- Seen in both humans and rodents
- Increased CCl2 levels associated with
  - Human Prostate cancer severity and progression
  - Human Gastric carcinomas
  - Human Oral carcinomas
  - Human Breast cancer
  - Human Thyroid cancer
  - Rat cholestatic liver injury
- May be related to PPAR signaling

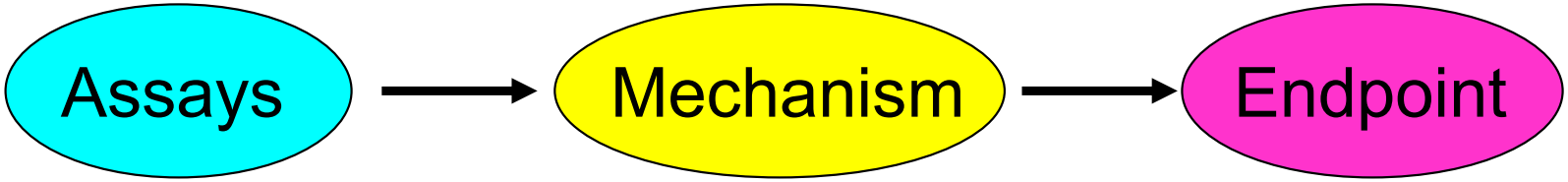
United States Environmental Protection Agency

# Rat Liver Disease Progression Links




Links Drawn for Univariate Associations with p<0.01

# Toxicity Signature Definition

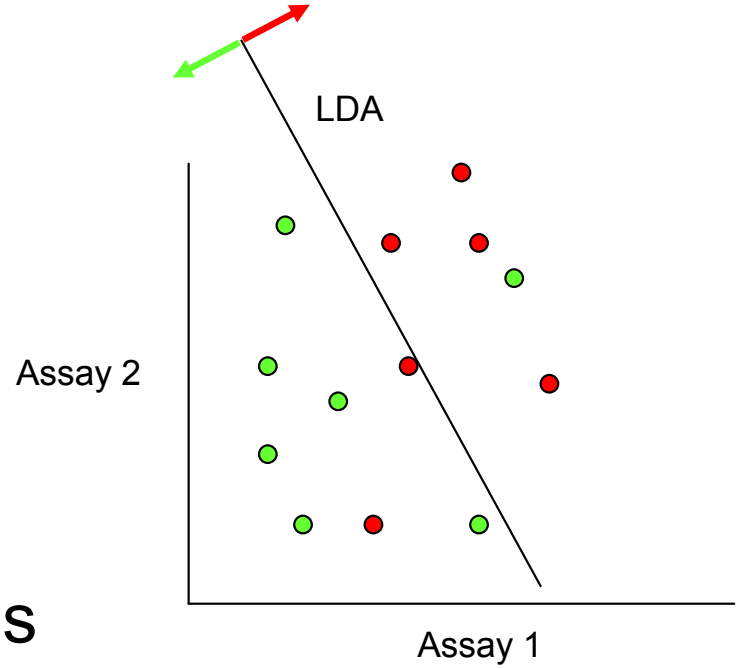



- An algorithm that takes as its input
  - A chemical
  - One or more *in vitro* assay measurement or *in silico* parameters
- And returns
  - A classification for that chemical for a toxicity endpoint

- Other terms
  - Model
  - Classifier

# Association Analysis / Signatures

- Use Machine Learning methods
  - SLR: Stepwise Logistic Regression
  - LDA: Linear Discriminant Analysis
  - SVM: Support Vector Machines
  - Many others
- For each binary endpoint, build models of form
  - *Predictor* = *F*(assay values)
  - If
    - *Predictor* for a chemical meets criteria
  - Then
    - Predict endpoint to be positive for the chemical




| | + Truth | - |
|---|---|---|
| + Test | TP | FP |
| - | FN | TN |

# Machine Learning Process

- ML Methods used
  - SVM – Support Vector Machines
  - NNET – Neural Networks
  - LDA – Linear Discriminant Analysis
  - SLR – Stepwise Logistic Regression

} Seemed to consistently overfit
Consistent with unbalanced data set

- Use AC50/LEC Data and log transform
- T-test Feature Selection
  - $p<0.1$ for cutoff
  - Accept maximum of n(chemical)/10 feature
- Use 5-fold cross validation
- Evaluate performance using balanced accuracy (BA)
  - BA=average of sensitivity and specificity

# SLR Signature: Rat Liver Proliferative Lesions

| Assay | Coefficient | Gene | Gene Name |
|---|---|---|---|
| Intercept | -2.86 | | |
| ATG_PPARg_TRANS | 0.298 | PPARG | peroxisome proliferator-activated receptor gamma |
| NVS_ADME_hCYP3A4 | 0.614 | CYP3A4 | cytochrome P450, family 3, subfamily A, polypeptide 4 |
| CLM_OxidativeStress_24hr | 0.403 | H2AFX | H2A histone family, member X (oxidative stress) |
| BSK_SM3C_MCP1_up | 0.331 | CCL2 | chemokine (C-C motif) ligand 2 |
| BSK_BE3C_IL1a_down | 0.389 | IL1A | interleukin 1, alpha |
| ATG_RORg_TRANS | 0.51 | RORC | RAR-related orphan receptor C |
| BSK_BE3C_tPA_up | 0.386 | PLAT | plasminogen activator, tissue |
| CLM_Hepat_Steatosis_24hr | 0.181 | | |
| ATG_PPARa_TRANS | 0.254 | PPARA | peroxisome proliferator-activated receptor alpha |
| CLM_MitoticArrest_24hr | -0.322 | | |
| CLM_p53Act_72hr | 0.28 | TP53 | tumor protein p53 |
| ATG_Sp1_CIS | 0.195 | SP1 | Sp1 transcription factor |
| ATG_NRF2_ARE_CIS | -0.171 | NFE2L2 | nuclear factor (erythroid-derived 2)-like 2 (oxidative stress) |

Start with 624 Assay measurements, 3 p-chem, 103 chemical structure class variables
Genes associated with tumors or liver disease in red

# Signature Performance – Proliferative Lesions

*In vivo* data

| Signature | + | - |
|---|---|---|
| + | 31 | 11 |
| - | 30 | 176 |

Sensitivity=51%
Specificity=94%

- 248/309 chemicals had rat data in ToxRefDB (used for model building)
- 8 other chemicals were predicted to be positive
  - PFOA: Causes rat liver adenomas
  - PFOS: Causes rat liver adenomas
  - Diniconazole: rat liver hypertrophy
  - Chlorothalonil: rat liver enlargement, kidney tumors
  - TCMTB: testicular and thyroid adenomas
  - No data for Niclosamide, Methylene bis(thiocyanate), Phenoxyethanol

United States Environmental Protection Agency

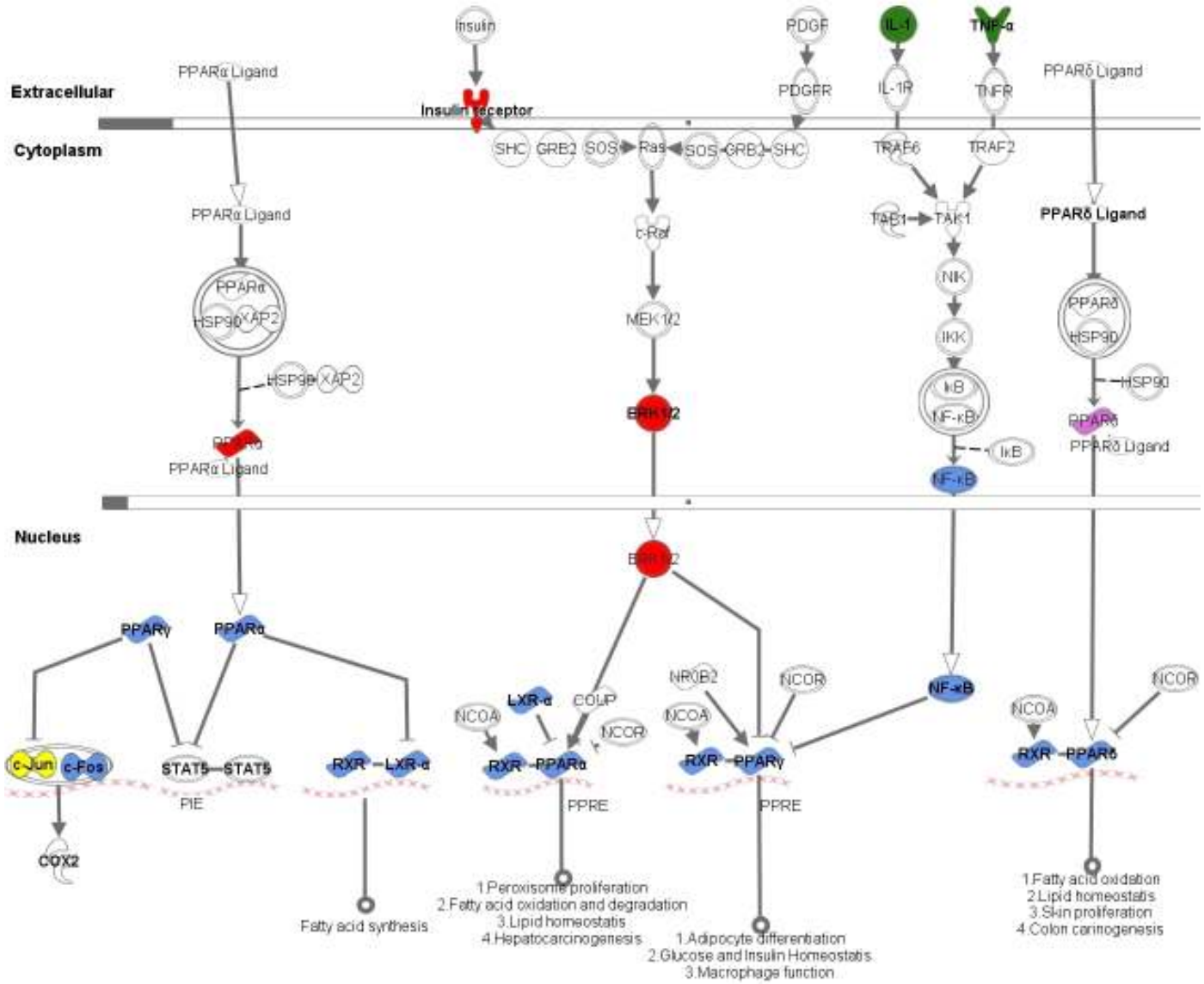
# Examine False Positives

- Look for data outside of ToxRefDB for highest scoring false positives

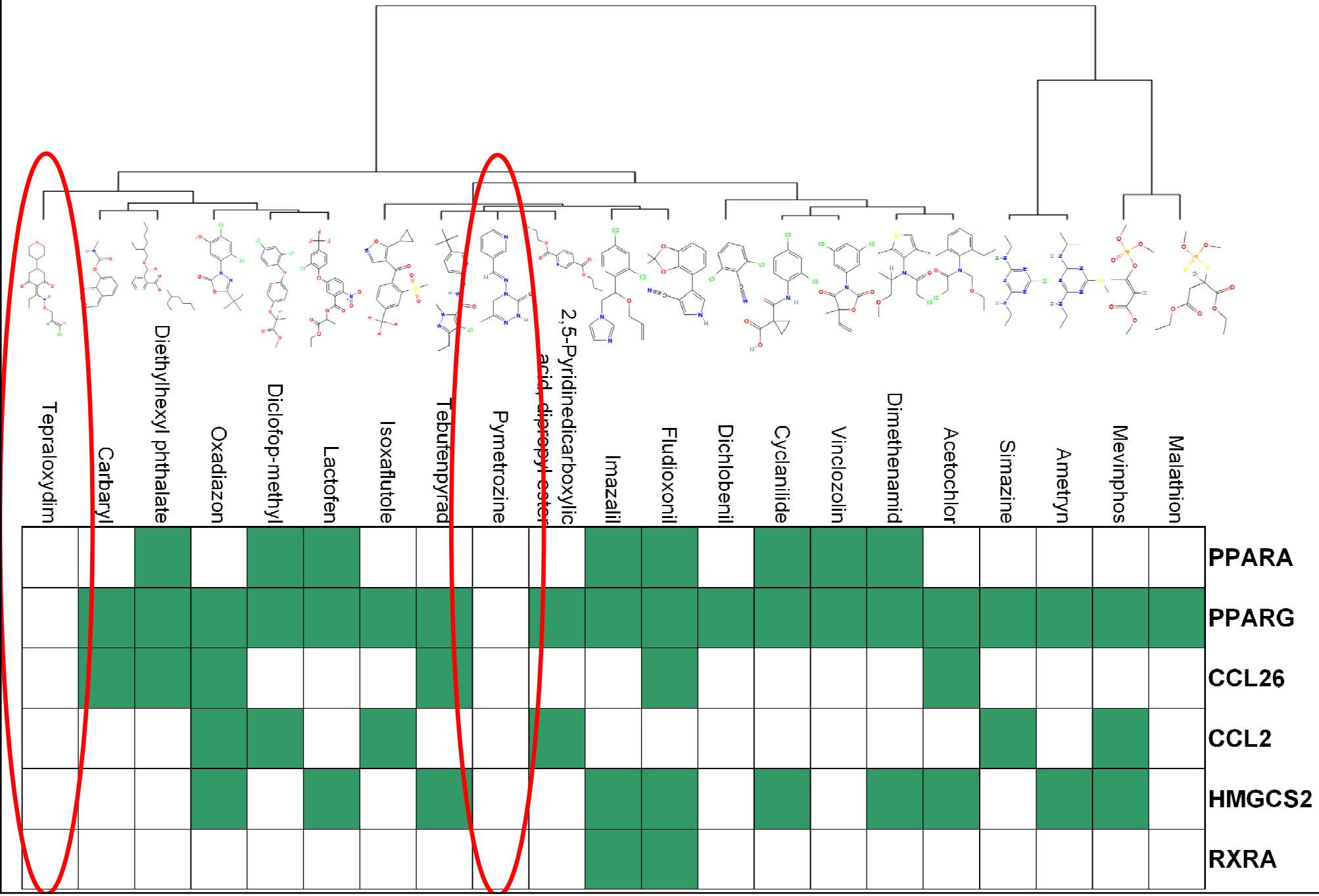
- Fenpyroximate
  - Liver hypertrophy in a rat 90-day subchronic study
- Bromoxynil
  - Non-proliferative lesions (2 year rat study)
  - Liver adenomas (2 year mouse study)
- Cyproconazole
  - Hepatocellular adenomas and carcinomas in mice
- Tribufos
  - Liver hemangiosarcomas in male mice

# Probing the PPAR Pathway – 33 Assays

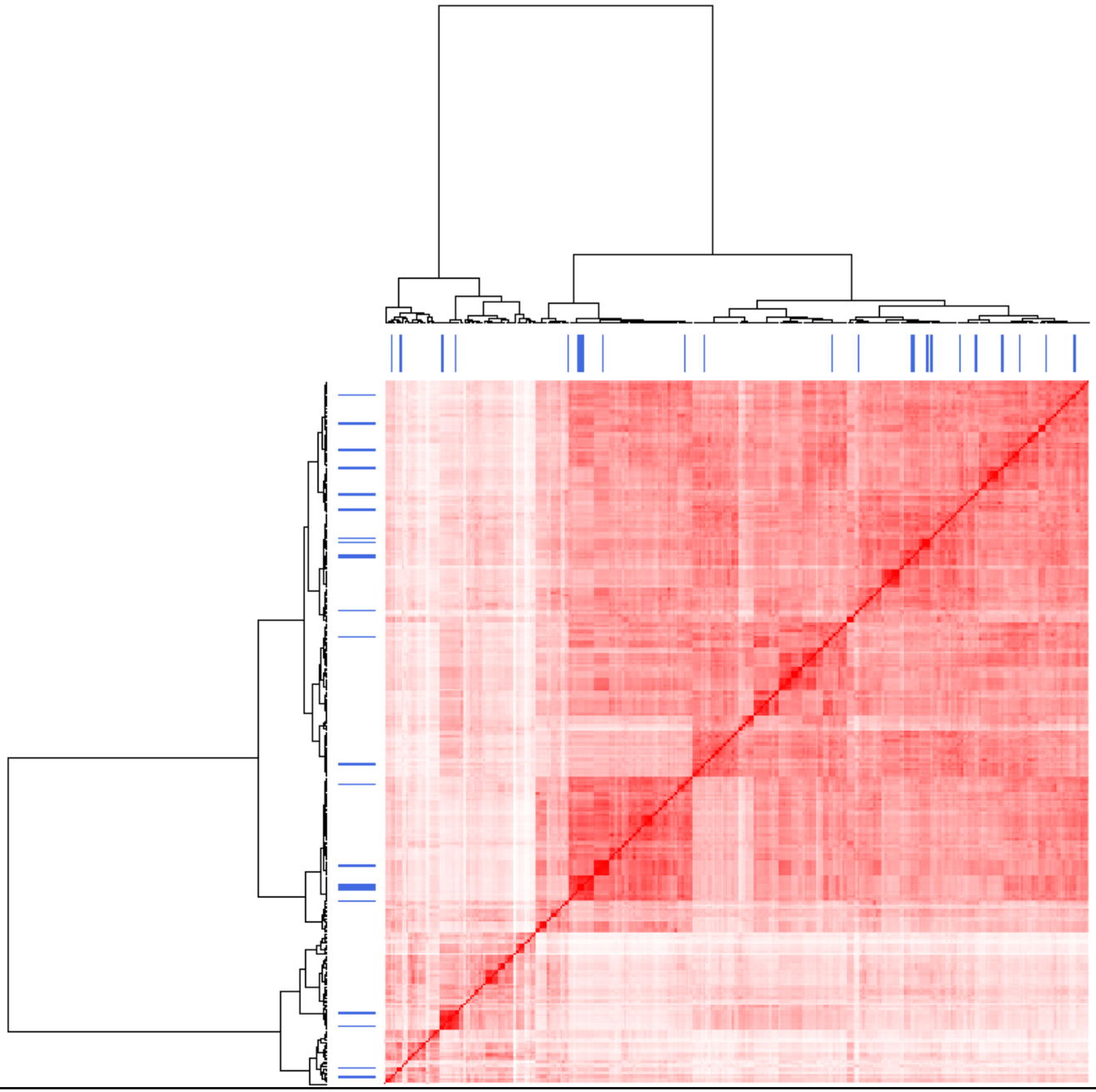
Chemicals causing liver proliferative lesions hit significantly increased number of assays in PPAR Pathway



# Rat Liver Tumorigens are diverse in chemical structure and *in vitro* Signature



# Rat Liver Tumorigens vs. Tanimoto Distance Clusters



Simple structure clustering does not bring liver tumorigens together

Will more sophisticated structure-based methods do better?

United States Environmental Protection Agency

# Open Research Issues

- Need better models / machine learning applications
  - The other talks should provide some answers
- Think about pathways
  - Mortensen et al. talk & several posters
- Think about biotransformation / PK
  - Martin et al. and Thomas et al. talks
- Bring in more assays
  - ToxCast and Tox21, several posters
- Add more chemicals, some with human tox data
  - Pfizer and HESI collaborations

# Toxicity is Multi-factorial